

FIGURE 1**Blastp vs. NCBI-nr**

```
>dbj|BAB68513.1| hatching enzyme EHE4 [Anguilla japonica]
Length = 271

Score = 197 bits (502), Expect = 1e-49
Identities = 103/233 (44%), Positives = 141/233 (60%), Gaps = 5/233 (2%)

Query: 52 DKDIPAINQGLILEETPESSFLIEGDIIRPSPFRLSATSNK--WPMGGSGVVEVPFLLS 109
        D D   I   ++      S   L+EGD+I   +   +   +N+   W   G+VEVP+   +S
Sbjct: 41 DPDDVDITTSILQSNNGSSEILMEGDLIVSNTRNAMKCWNQCLWRKSSDGLVEVPYTVS 100

Query: 110 SKYDEPSHQVILEALAEFERSTCIRFVTYDQRDFISIIPMYCCFSSVGRSGGMQVVSLA 169
        +++   + I A+ F TCIRFV   QRDFISI   GC+S +GR+GG QVVSLA
Sbjct: 101 NEFSYYHKKRIENAMKTFNTETCIRFVPRSSQRDFISIESRDGCYSYLGRTGGKQVVSLA 160

Query: 170 PT-CLQKGRGIVLHELMHVLGFWHEHTRADRDRYIRVNWNEILPGFEINFIKSQSSNMLT 228
        C+   GI+ HEL H LGF+HEHTR+DRD Y+++NW + P   NF   ++N+ T
Sbjct: 161 RYGCVY--HGIIQHELNHALGFYHEHTRSDRDEYVKINWENVAPHTIYNFQTQDTNNLNT 218

Query: 229 PYDYSSVMHYGRLAFSRRGLPTITPLWAPSvhIGQRWNLSASDITRVLKLYGC 281
        PYDY+S+MHYGR AFS G+ TITP+ P+ IGQR ++S DI R+ KLY C
Sbjct: 219 PYDYTSIMHYGRTAFSTNGMDTITPVNPNPNSIGQRRSMSRGDILRIKKLYSC 271
```

Tblastn vs. NCBI-est

```
Tissue - Uterus tumour

>gb|BI061462.1|BI061462 IL3-UT0117-070301-494-H12 UT0117 Homo sapiens cDNA.
Length = 652

Score = 175 bits (443), Expect = 2e-42
Identities = 85/86 (98%), Positives = 85/86 (98%)
Frame = -2

Query: 29 SCAGACGTSPDGLTPEGTQASGDKDIPAINQGLILEETPESSFLIEGDIIRPSPFRLS 88
        SCAGACGTSPDGLTPEGTQASGDKDIPAINQGLILEETPESSFLIEGDIIRPSPFRLS
Sbjct: 546 SCAGACGTSPDGLTPEGTQASGDKDIPAINQGLILEETPESSFLIEGDIIRPSPFRLS 367

Query: 89 ATSNKWPMSGGVVEVPFLLSSKYDE 114
        ATSNKWPMSGGVVEVPFLLSSKY E
Sbjct: 366 ATSNKWPMSGGVVEVPFLLSSKYGE 289
```

FIGURE 2

Library	Tissue/cell source	Vector	Host strain	Supplier	Cat. no.
1	human fetal brain	Zap II	XLI-Blue MRF	Stratagene	936206
2	human ovary	GT10	LE392	Clontech	HL1098a
3	human pituitary	GT10	LE392	Clontech	HL1097a
4	human placenta	GT11	LE392	Clontech	HL1075b
5	human testis	GT11	LE392	Clontech	HL1010b
6	human substantia nigra	GT10	LE392	in house	
7	human fetal brain	GT10	LE392	in house	
8	human cortex brain	GT10	LE392	in house	
9	human colon	GT10	LE392	Clontech	HL1034a
10	human fetal brain	GT10	LE392	Clontech	HL1065a
11	human fetal lung	GT10	LE392	Clontech	HL1072a
12	human fetal kidney	GT10	LE392	Clontech	HL1071a
13	human fetal liver	GT10	LE392	Clontech	HL1064a
14	human bone marrow	GT10	LE392	Clontech	HL1058a
15	human peripheral blood monocytes	GT10	LE392	Clontech	HL1050a
16	human placenta	GT10	LE392	in house	
17	human SHSYSY	GT10	LE392	in house	
18	human U373 cell line	GT10	LE392	in house	
19	human CFPoc-1 cell line	Uni Zap	XLI-Blue MRF	Stratagene	936206
20	human retina	GT10	LE392	Clontech	HL1132a
21	human urinary bladder	GT10	LE392	in house	
22	human platelets	Uni Zap	XLI-Blue MRF	in house	
23	human neuroblastoma Kan + TS	GT10	LE392	in house	
24	human bronchial smooth muscle	GT10	LE392	in house	
25	human bronchial smooth muscle	GT10	LE392	in house	
26	human Thymus	GT10	LE392	Clontech	HL1127a
27	human spleen 5' stretch	GT11	LE392	Clontech	HL1134b
28	human peripheral blood monocytes	GT10	LE392	Clontech	HL1050a
29	human testis	GT10	LE392	Clontech	HL1065a
30	human fetal brain	GT10	LE392	Clontech	HL1065a
31	human substantia nigra	GT10	LE392	Clontech	HL1093a
32	human placenta#11	GT11	LE392	Clontech	HL1075b
33	human Fetal brain	GT10	LE392	Clontech	custom
34	human placenta #59	GT10	LE392	Clontech	HLS014a
35	human pituitary	GT10	LE392	Clontech	HL1097a
36	human pancreas #63	Uni Zap XR	XLI-Blue MRF	Stratagene	937208
37	human placenta #19	GT11	LE392	Clontech	HL1008
38	human liver 5'stretch	GT11	LE392	Clontech	HL1115b
39	human uterus	Zap-CMV XR	XLI-Blue MRF	Stratagene	980207
40	human kidney large-insert cDNA library	TriplEx2	XLI-Blue	Clontech	HL5507u

FIGURE 3

1 AGGTCTTGT GGACAAATAGC TATTCTTCTT GGCTCTGTCG CTTCCCTCA CTGGGTGCAG
 61 GTGACTGTGG GGGTGTCccc AAATGCTGCC CAGCGCTGAC ATGCTCCGCC TCTGGGATTT
 m l r 1 w d
 121 CAATCCAGGT GGGGCCCTGA GTGACCTGGC TCTGGGGCTC AGGGGTATGG AGGAGGGGGG
 f n p g g a l s d l a l g l r g m e e g
 181 ATATAGCTGC GCAGGGAGCCT GTGGTACCAAG CTTCCCAGAT GCCCTCACCC CTGAGGGAAC
 g y s c a g a c g t s f p d g l t p e g
 241 CCAGGCCTCC GGGGACAAGG ACATTCTGC AATTAACCAA GGGCTCATCC TGGAAGAAC
 t q a s g d k d i p a i n q g l i l e e
 301 CCCAGAGAGC AGCTTCCTCA TCGAGGGGGA CATCATCCGG CCGAGTCCCT TCCGACTGCT
 t p e s s f l i e g d i i r p s . p f r l
 361 GTCAGCAACC AGCAACAAAT GGCCCATGGG TGGTAGTGGT GTCGTGGAGG TCCCCTTCTC
 l s a t s n k w p m g g s g v v e v p f
 421 GCTCTCCAGC AAGTACGATG AGCCCAGCCA TCAGGTCATC CTGGAGGCTC TTGCGGAGTT
 l l s s k y d e p s h q v i l e a l a e
 481 TGAACGTTCC ACGTGCATCA GGTTTGTAC CTATCAGGAC CAGAGAGACT TCATTTCCAT
 f e r s t c i r f v t y q d q r d f i s
 541 CATCCCCATG TATGGGTGCT TCTCGAGTGT GGGGCGCAGT GGAGGGATGC AGGTGGTCTC
 i i p m y g c f s s v g r s g g m q v v
 601 CCTGGCGCCC ACGTGTCTCC AGAAGGGCCG GGGCATTGTC CTTCATGAGC TCATGCATGT
 s l a p t c l q k g r g i v l h e l m h
 CP1
 661 GCTGGCTTC TGGCACGAGC ACACGGGGC CGACCGGGAC CGCTATATCC GTGTCAACTG
 v l g f w h e h t r a d r d r y i r v n
 721 GAACGAGATC CTGCCAGGCT TTGAAATCAA CTTCATCAAG TCTCAGAGCA GCAACATGCT
 w n e i l p g f e i n f i k s q s s n m
 781 GACGCCCTAT GACTACTCCT CTGTGATGCA CTATGGGAGG CTCGCCCTCA GCCGGCGTGG
 l t p y d y s s v m h y g r l a f s r r
 78836-GR1-3'
 841 GCTGCCACC ATCACACCAAC TTTGGGCCCC CAGTGTCCAC ATCGGCCAGC GATGGAACCT
 g l p t i t p l w a p s v h i g q r w n

 901 GAGTGCCTCG GACATCACCC GGGCCTCAA ACTCTACGGC TGCAGCCCAA GTGGCCCCAG
 l s a s d i t r v l k l y g c s p s g p
 78836-GR1nest-3'
 961 GCCCCGTGGG AGAGGTGAGT GGCATGGCAG GAAGGTGACT TGAAACCTGGA GAAGGCGCCT
 r p r g r g e w h g r k v t -
 1021 GTGCTCTAAT GGTGTCAGGG AGGGTGACAA GGAGGGAGAT GAGGTTGCAG GGGGAGCAGG
 1081 GTGAGATCAC GGGGGCTTGC CAC

Position and sense of PCR primers

FIGURE 4

Primer	Name	Sequence (5'-3')
CP1	4C5	ACC GCT ATA TCC GTG TCA A
CP2	4C6	GCT GCA GCC GTA GAG TTT
GeneRacer 3'		GCT GTC AAC GAT ACG CTA CGT AAC G
78836-GR1-3'		AGT GTC CAC ATC GGC CAG CGA TGG AA
GeneRacer 3' nested		CGC TAC GTA ACG GCA TGA CAG TG
78836-GR1nest-3'		ATG GAA CCT GAG TGC CTC GGA CAT C
78836-FL-F	4C7	CTG TCA GCA ACC AGC AAC AA
78836-FL-R	9B2	AGC CAC AGG CTT AAT CTT CG
78836-FL2-F	9E6	TCT ACC ATG GAG GGT GTA GG

FIGURE 5

```

1 ATGGAACCTG AGTGCCTCGG ACATCACCCG GGTCTCTAAA CTCTACGGCT GCAGCCCAAG
w n l s a s d i t r v l k l y g c s p

61 TGGCCCCAGG CCCC GTGGGA GAGGGTCCCA TGCCCACAGC ACTGGTAGGA GCCCGCCTCC
s g p r p r g r g s h a h s t g r s p a

121 GGCCTCCCTA TCTCTGCAGC GGCTTTGGA GGCACTGTGCG GCGGAATCCA GGAGCCCCGA
p a s l s l q r l l e a l s a e s r s p

181 CCCCAGTGGT TCCAGTGCAGG GAGGCCAGCC CGTTCTGCA GGGCCTGGGG AGAGCCACA
d p s g s s a g g q p v p a g p g e s p

241 TGGGTGGGAG TCCCCTGCC TGAAAAAGCT CAGTGCAGAG GCCTCGGCAA GGCAGCCTCA
h g w e s p a l k k l s a e a s a r q p

301 GACCCTAGCT TCCTCCCCAA GATCAAGGCC TGGAGCAGGT GCCCCCCGGTG TTGCTCAGGA
q t l a s s p r s r p g a g a p g v a q

361 GCAGTCCTGG CTGGCCGGAG TGTCCACCAA GCCCACAGTC CCATCTTCAG AAGCAGGAAT
e q s w l a g v s t k p t v p s s e a g

421 CCAGCCAGTC CCTGTCCAGG GAAGCCCAGC TCTGCCAGGG GGCTGTGTAC CTAGAAATCA
i q p v p v q g s p a l p g g c v p r n

481 TTTCAAGGGG ATGTCCGAAG ATTAAAGCCTG TGGCTTCTGT CCCCAAGTAG GGAGGGCATC
h f k g m s e d

541 CTCTGCCAG TGGAGCTGGG TCGTCTACCT CTTGGCTCCT TTGGGCCACA CCACTGTCTT
601 CCAGCCCCAA CCTACCAACCC CATCTCAGAG GGCCAGGACT CTTCCCTGT CTCTCTTCAC
661 TGTGTTCCCC TAAGGGCTCC TAGGGCCAGG GGTCTTCTTA GCTCTGCCAC AGGGGAAGGC
721 AGGCCTGGCT GTGCCTGCTC TTGACTTTG CCCAGCCCTG GTGGATGCTG GGAATGGGAG
781 GTGACATTCT CCAGGGACAG GTCCTGGAAG GGGTGGGAA GAGGTAGGTT CCAGCCCCGC
841 AGAACCCCTGG AATCCCTCCT GTGCCTGAGG CCCTGCCCTC CAGCATGGAC TAATGGTGTG
901 CCTACCTCTC CCTCAGGGCA GCCCTGTGGC TGGGACCCCTG GGAACAGCCT CCCATCCAC
961 CCAACATGCC CAAGTGTGGG GGAATGTTCT ACAGCAGTGT AGCCTCCAGC CCTTCTCTCC
1021 AGGAGGCTTT GAGAGCCCAA CTTACTCCCC TGCAAGCAG GAAGGTGGTA GGTCAAGTGT
1081 GGCCACCATG GGGGAGACGA GAAAGAAGTG GGGCCCCACC AGATTGCACA ATGGGAACCT
1141 CAGCTGGCCC CTGAACAGAG GACTCAGTTG TCTCCACCT ACACCGCTAT TCCCTGGAGC
1201 TCAGCCAGGC GCAGCCTTGG AAGGAGAAAG GGCTGGGGTT ACCTGGCTTG TCCTCCCTCA
1261 GGAAAGCCCC CTTCTCTCTC TGCCCCAGCT CCCAGCCTGG CCTCTCTCCAG GCAGGCCCTA
1321 CTCTCTGCC CCAGCTCCGG CTTTCCCCAT GAGGTTTGTG CCAGGCATGA AGAAAGCATC
1381 CAGGGTGCCA ATGAGTGGGC CTAGGCCAGA GGCCCCCTCAG TCCCCAAGGG TACTGTTTG
1441 GTGGCCTTTC AGAGGGTCAA GGAAGCCCTG CTTGGGGTAG AAGGGGCAGG AGCCCCACAT
1501 GTTGGGGGAG GAAATAAAAGT GGAGTGTGCT GTGCTGAAAA AAAAAAAA AAAA

```

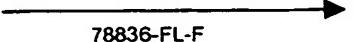
TAA Stop codon

AATAAA Consensus polyadenylation site

FIGURE 6

Primer	Sequence (5'-3')
T3	ATT AAC CCT CAC TAA AGG GA
T7	TAA TAC GAC TCA CTA TAG GG
SP6	ATT TAG GTG ACA CTA TAG

FIGURE 7

1 CTGTCAGCAA CCAGCAACAA ATGGCCCATG GGTGGTAGTG GTGTCGTGGA GGTCCCCCTTC


61 CTGCTCTCCA GCAAGTACGA TGAGCCCAGC CGCCAGGTCA TCCTGGAGGC TCTTGCAGGAG
 1 l s s k y d e p s r q v i l e a l a e

121 TTTGAACGTT CCACGTGCAT CAGGTTTGTC ACCTATCAGG ACCAGAGAGA CTTCATTTCC
 f e r s t c i r f v t y q d q r d f i s

181 ATCATCCCCA TGTATGGGTG CTTCTCGAGT GTGGGGCGCA GTGGAGGGAT GCAGGTGGTC
 i i p m y g c f s s v g r s g g m q v v

241 TCCCTGGCGC CCACGTGTCT CCAGAAGGGC CGGGGCATTG TCCTTCATGA GCTCATGCAT
 s l a p t c l q k g r g i v l h e l m h

301 GTGCTGGGCT TCTGGCACGA GCACACGCGG GCCGACCAGG ACCGCTATAT CCGTGTCAAC
 v l g f w h e h t r a d r d r y i r v n

361 TGGAACGAGA TCCTGCCAGG CTTTGAAATC AACTTCATCA AGTCTCAGAG CAGAACATG
 w n e i l p g f e i n f i k s q s s n m

421 CTGACGCCCT ATGACTACTC CTCTGTGATG CACTATGGGA GGCTCGCCTT CAGCCGGCGT
 l t p y d y s s v m h y g r l a f s r r

481 GGGCTGCCCA CCATCACACC ACTTTGGGCC CCCAGTGTCC ACATCGGCCA GCGATGGAAC
 g l p t i t p l w a p s v h i g q r w n

541 CTGAGTGCCT CGGACATCAC CGGGGTCTC AAACTCTACG GCTGCAGCCC AAGTGGCCCG
 l s a s d i t r v l k l y g c s p s g p

601 AGGCCCGTG GGAGAGGGTC CCATGCCAC AGCACTGGTA GGAGCCCCGC CCCGGCTCC
 r p r g r g s h a h s t g r s p a p a s

661 CTATCTCTGC AGCGGCTTT GGAGGCCTG TCGGGCGGAAT CCAGGAGGCC CGACCCAGT
 l s l q r l l e a l s a e s r s p d p s

721 GGTTCCAGTG CGGGAGGCCA GCCCGTTCT GCAGGGCTTG GGGAGAGCCC ACATGGGTGG
 g s s a g g q p v p a g p g e s p h g w

781 GAGTCCCTG CCCTGAAAAA GCTCAGTGCA GAGGCCTCGG CAAGGCAGCC TCAGACCCCTA
 e s p a l k k l s a e a s a r q p q t l

841 GCTTCCTCCCC CAAGATCAAG GCCTGGAGCA GGTGCCCG GTGTTGCTCA GGAGCAGTCC
 a s s p r s r p g a g a p g v a q e q s

901 TGGCTGGCCG GAGTGTCCAC CAAGCCCCACA GTCCCATCTT CAGAAGCAGG AATCCAGCCA
 w l a g v s t k p t v p s s e a g i q p

961 GTCCCTGTCC AGGGAGGCC AGCTCTGCCA GGGGGCTGTG TACCTAGAAA TCATTTCAAG
 v p v q g s p a l p g g c v p r n h f k

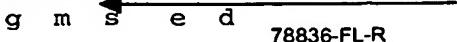
1021 GGGATGTCCG AAGATTAAGC CTGTGGCT


FIGURE 8

Query= INSP005a
(336 letters)

Database: All non-redundant GenBank CDS
translations+PDB+SwissProt+PIR+PRF
1,247,039 sequences; 397,579,747 total letters

Searching.....done

		Score	E
		(bits)	Value
Sequences producing significant alignments:			
ref XP_141346.1	similar to hatching enzyme EHE7 [Anguilla japonica]	416	e-115
dbj BAB68518.1	hatching enzyme EHE13 [Anguilla japonica]	187	2e-46
dbj BAB68515.1	hatching enzyme EHE7 [Anguilla japonica]	186	4e-46
dbj BAB68516.1	hatching enzyme EHE10 [Anguilla japonica]	186	4e-46
dbj BAB68513.1	hatching enzyme EHE4 [Anguilla japonica]	186	5e-46
dbj BAB68517.1	hatching enzyme EHE12 [Anguilla japonica]	183	3e-45
dbj BAB68514.1	hatching enzyme EHE6 [Anguilla japonica]	183	3e-45
dbj BAB68519.1	hatching enzyme EHE14 [Anguilla japonica]	182	7e-45
pir C48826	high choriolytic hatching proteinase (EC 3.4.24.-) H...	171	1e-41
dbj BAA12146.1	choriolyisin H [Oryzias latipes]	171	2e-41

Top alignment to known metalloproteinase:

```
>dbj|BAB68518.1| hatching enzyme EHE13 [Anguilla japonica]
Length = 271

Score = 187 bits (475), Expect = 2e-46
Identities = 93/183 (50%), Positives = 124/183 (66%), Gaps = 3/183 (1%)

Query: 5   GVVEVPFLLSSKYDEPSRQVILEALAEFERSTCIRFVTYQDQRDFISIIPMYGCFSSVGR 64
          G+VEVP+ +SS++    ++ I  A+ F   TCIRFV   QRDFISI   GC+S +GR
Sbjct: 91  GLVEVPYPTVSSEFSYYHKKRIENAMETFNTETCIRFVPRSSQRDFISIESRDGCYSYLGR 150

Query: 65  SGGMQVVSLAPT-CLQKGRGIVLHELMHVLGFWHEHTRADRDRYIRVNNEILPGFEINF 123
          +GG QVVSLA   C+   GI+ HEL H LGF+HEHTR+DRD Y+++NW   + P   NF
Sbjct: 151 TGGKQVVSLARYGCVY--HGIIQHELNHALGFYHEHTRSDRDEYVKINWENVAPHTIYNF 208

Query: 124 IKSQSSNMLTPYDYSSVMHYGRLAFSRRGLPTITPLWAPSvhIGQRWNLSASDITRVLKL 183
          +  ++N+ TPYDY+S+MHYGR AFS  G+ TITP+  P+  IGQR  ++S  DI R+ KL
Sbjct: 209 QEQQDTNNLNTPYDYTSIMHYGRTAFSTNGMDTITPVPNPNQSIGQRRSMSKGDIILRINKL 268

Query: 184 YGC 186
          Y C
Sbjct: 269 YSC 271
```

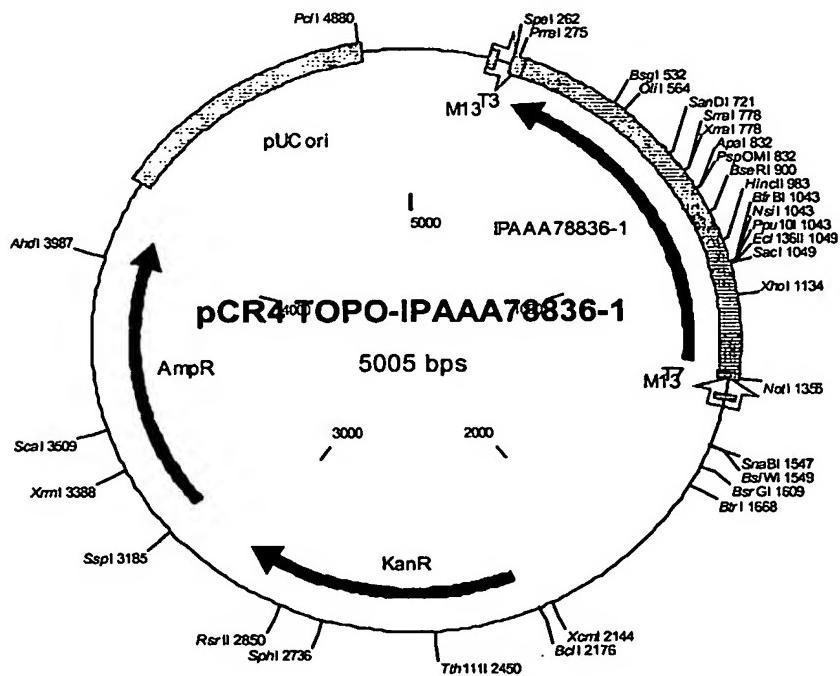
FIGURE 9

Molecule: pCR4 TOPO-IPAAA78836-1, 5005 bps DNA Circular
 File Name: 13164.cm5, dated 24 Oct 2002

Description: Ligation of inverted 78836_F2/R8 PCR product into pCR4-TOPO linear vector*

Molecule Features:

Type	Start	End	Name	Description
REGION	205	221	M13	rev priming site
MARKER	243		T3	
REGION	262	294		'Polylinker'
REGION	294	294		'TOPO cloning site'
GENE	1315	308	C IPAAA78836-1	
REGION	1342	295	C	Inserted PCR product
REGION	1343	1360		'Polylinker'
REGION	1343	1343		'TOPO cloning site'
MARKER	1395		C T7	
REGION	1403	1418	M13	
GENE	2207	3001	KanR	
GENE	3205	4065	AmpR	
REGION	4210	4883	pUC ori	



10/18

FIGURE 10

78836-FL2-F

```

1 TTCTACCATG GAGGGTGTAG GGGGTCTCTG GCCTTGGGTG CTGGGTCTGC TCTCCTTGCC
   m e g v g g l w p w v l g 1 1 s 1

61 AGGTGTGATC CTAGGAGCGC CCCTGGCCTC CAGCTGCAGCA GGAGCCTGTG GTACCAGCTT
   p g v i l g a p l a s s c a g a c g t s

121 CCCAGATGGC CTCACCCCTG AGGAAACCCA GGCCTCCGGG GACAAGGACA TTCCTGCAAT
   f p d g l t p e g t q a s g d k d i p a

181 TAACCAAGGG CTCATCCTGG AAGAAACCCC AGAGAGCAGC TTCCATCG AGGGGGACAT
   i n q g l i l e e t p e s s f l i e g d

241 CATCCGGCCG AGTCCCTTCC GACTGCTGTC AGCAACCAGC AACAAATGGC CCATGGGTGG
   i i r p s p f r l l s a t s n k w p m g

301 TAGTGGTGTC GTGGAGGTCC CCTTCCTGCT CTCCAGCAAG TACGATGAGC CCAGCCGCCA
   g s g v v e v p f l l s s k y d e p s r

361 GGTCACTCTG GAGGCTCTTG CGGAGTTGA ACGTTCCACG TGCACTCAGGT TTGTCACCTA
   q v i l e a l a e f e r s t c i r f v t

421 TCAGGACCAAG AGAGACTTCA TTTCCATCAT CCCATGTAT GGGTGCTTCT CGAGTGTGGG
   y q d q r d f i s i i p m y g c f s s v

481 GCGCAGTGGA GGGATGCAGG TGGTCTCCCT GGCGCCACG TGTCTCCAGA AGGGCCGGGG
   g r s g g m q v v s l a p t c l q k g r

541 CATTGTCTT CATGAGCTCA TGCATGTGCT GGGCTTCTGG CACGAGCACA CGCGGGCCGA
   g i v l h e l m h v l g f w h e h t r a

601 CCGGGACCGC TATATCCGTG TCAACTGGAA CGAGATCCTG CCAGGCTTTG AAATCAACTT
   d r d r y i r v n w n e i l p g f e i n

661 CATCAAGTCT CGGAGCAGCA ACATGCTGAC GCCCTATGAC TACTCCTCTG TGATGCACTA
   f i k s r s s n m l t p y d y s s v m h

721 TGGGAGGCTC GCCTTCAGCC GGCCTGGCT GCCCACCATC ACACCACTTT GGGCCCCAG
   y g r l a f s r r g l p t i t p l w a p

781 TGTCCACATC GGCCAGCGAT GGAACCTGAG TGCCTCGGAC ATCACCCGGG TCCTCAAAC
   s v h i g q r w n l s a s d i t r v l k

841 CTACGGCTGC AGCCCAAGTG GCCCCAGGCC CCGTGGGAGA GGGTCCCATG CCCACAGCAC
   l y g c s p s g p r p r g r g s h a h s

901 TGGTAGGAGC CCCGCTCCGG CCTCCCTATC TCTGCAGCGG CTTTTGGAGG CACTGTCGGC
   t g r s p a p a s l s l q r l l e a l s

961 GGAATCCAGG AGCCCCGACC CCAGTGGTTC CAGTGCAGGA GGCCAGCCCG TTCCCTGCAGG
   a e s r s p d p s g s s a g g q p v p a

1021 GCCTGGGGAG AGCCACATG GGTGGGAGTC CCCTGCCCTG AAAAAGCTCA GTGCAGAGGC
   g p g e s p h g w e s p a l k k l s a e

```

11/18

1081 CTCGGCAAGG CAGCCTCAGA CCCTAGCTTC CTCCCCAAGA TCAAGGCCTG GAGCAGGTGC
a s a r q p q t l a s s p r s r p g a g
1141 CCCCGGTGTT GCTCAGGAGC AGTCCTGGCT GGCCGGAGTG TCCACCCAAGC CCACAGTCCC
a p g v a q e q s w l a g v s t k p t v
1201 ATCTTCAGAA GCAGGAATCC AGCCAGTCCC TGTCCAGGGGA AGCCCAGCTC TGCCAGGGGG
p s s e a g i q p v p v q g s p a l p g
1261 CTGTGTACCT AGAAAATCATT TCAAGGGGAT GTCCGAAGAT TAAGCCTGTG GCT
g c v p r n h f k g m s e d

78836-FL-R

12/18

FIGURE 11

Query= INSP005b
 (431 letters)

Database: All non-redundant GenBank CDS
 translations+PDB+SwissProt+PIR+PRF
 1,247,039 sequences; 397,579,747 total letters

Searching.....done

Sequences producing significant alignments:	Score (bits)	E Value
ref XP_141346.1 similar to hatching enzyme EHE7 [Anguilla japonica]	540	e-152
dbj BAB68513.1 hatching enzyme EHE4 [Anguilla japonica]	198	1e-49
dbj BAB68518.1 hatching enzyme EHE13 [Anguilla japonica]	198	1e-49
dbj BAB68516.1 hatching enzyme EHE10 [Anguilla japonica]	197	3e-49
dbj BAB68515.1 hatching enzyme EHE7 [Anguilla japonica]	196	4e-49
dbj BAB68514.1 hatching enzyme EHE6 [Anguilla japonica]	196	7e-49
dbj BAB68517.1 hatching enzyme EHE12 [Anguilla japonica]	194	3e-48
dbj BAB68519.1 hatching enzyme EHE14 [Anguilla japonica]	191	1e-47
pir C48826 high choriolytic hatching proteinase (EC 3.4.24.-) H...	187	3e-46
dbj BAA12146.1 choriolysin H [Oryzias latipes]	186	4e-46

Top alignment to known metalloproteinase:

>dbj|BAB68518.1| hatching enzyme EHE13 [Anguilla japonica]
 Length = 271

Score = 198 bits (503), Expect = 1e-49
 Identities = 103/233 (44%), Positives = 144/233 (61%), Gaps = 5/233 (2%)

Query: 52 DKDIPAINQGLILEETPESSFLIEGDIIRPSPFRLLSATSNK--WPMGGSGVVEVPFLLS 109
 D D I ++ S L+EGD++ + ++ +N+ W G+VEVP+ +S
 Sbjct: 41 DPDDLDITARILQSNNGSSEILMEGDMVVSNTRNAINCWNQNCLWRKSSDGLVEVPYTVS 100

Query: 110 SKYDEPSRQVILEALAEFERSTCIRFVTYQDQRDFISIIPMYGCFSSVGRSGGMQVVSLA 169
 S++ ++ I A+ F TCIRFV QRDFISI GC+S +GR+GG QVVSLA
 Sbjct: 101 SEFSYYHKKRIENAMETFNTETCIRFVPRSSQRDFISIESRDGCYSYLGRTGGKQVVSLA 160

Query: 170 PT-CLQKGRGIVLHELMHVLGFWHEHTRADRDYIRVNWNEILPGFEINFIKSRSNNMLT 228
 C+ GI+ HEL H LGF+HEHTR+DRD Y+++NW + P NF + ++N+ T
 Sbjct: 161 RYCCVY--HGIIQHELNHALGFYHEHTRSDRDEYVKINWENVAPHTIYNFQEQTNNNLNT 218

Query: 229 PYDYSSVMHYGRLAFSRRGLPTITPLWAPSvhIGQRWNLSASDITRVLKLYGC 281
 PYDY+S+MHYGR AFS G+ TITP+ P+ IGQR ++S DI R+ KLY C
 Sbjct: 219 PYDYTSIMHYGRTAFSTNGMDTITPPNPQNSIGQRRSMSKGDIILRINKLYSC.

FIGURE 12

Molecule: pCR4 TOPO-IPAAA78836-2, 5269 bps DNA Circular
 File Name: 13296.cm5, dated 24 Oct 2002

Description: Ligation of inverted IPAAA78836v2 into pCR4-TOPO linear vector*

Molecule Features:

Type	Start	End	Name	Description
REGION	205	221	M13	rev priming site
MARKER	243		T3	
REGION	262	294		Polylinker'
REGION	294	294		TOPO cloning site'
GENE	1600	307	C IPAAA78836-2	
REGION	1607	1624		'Polylinker
REGION	1607	1607		'TOPO cloning site
MARKER	1659		C T7	
REGION	1667	1682	M13	
GENE	2471	3265	KanR	
GENE	3469	4329	AmpR	
REGION	4474	5147	pUC ori	

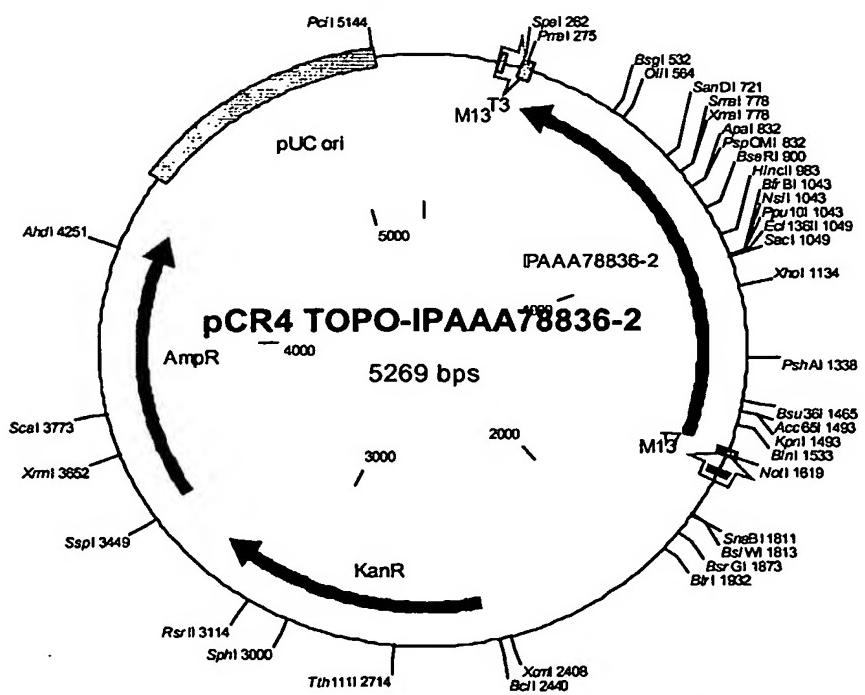


FIGURE 13

Active site residues are highlighted in grey below.

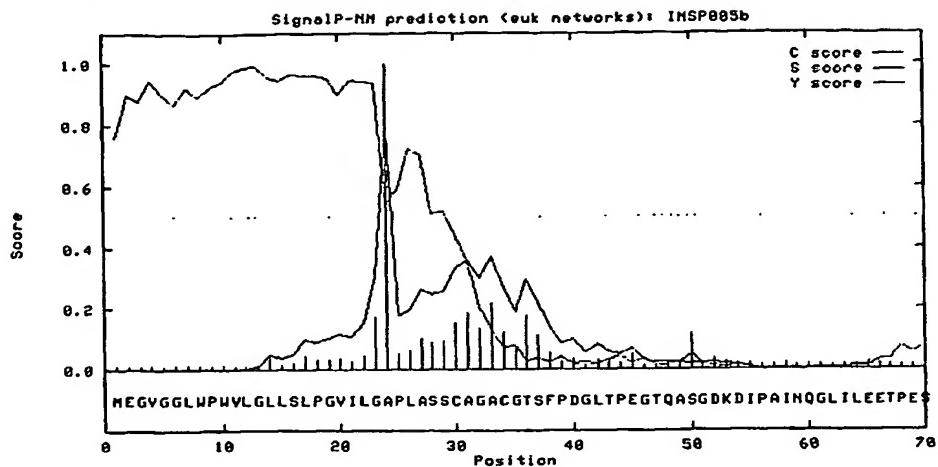
WO2002/16566-A2	-----
AX526191	AEASARQPQTTLASSPRSRPGAGAPGVAQEQSFLAGVSTKPTVPSSEAGIQPVPVQGSPAL
INSP005 PREDICTION	-----
INSP005b	AEASARQPQTTLASSPRSRPGAGAPGVAQEQSFLAGVSTKPTVPSSEAGIQPVPVQGSPAL
INSP005a	AEASARQPQTTLASSPRSRPGAGAPGVAQEQSFLAGVSTKPTVPSSEAGIQPVPVQGSPAL

WO2002/16566-A2 -----
AX526191 PGGCVPRNHFKGMSED
INSP005 PREDICTION -----
INSP005b PGGCVPRNHFKGMSED
INSP005a PGGCVPRNHFKGMSED

FIGURE 14

>INSP005b

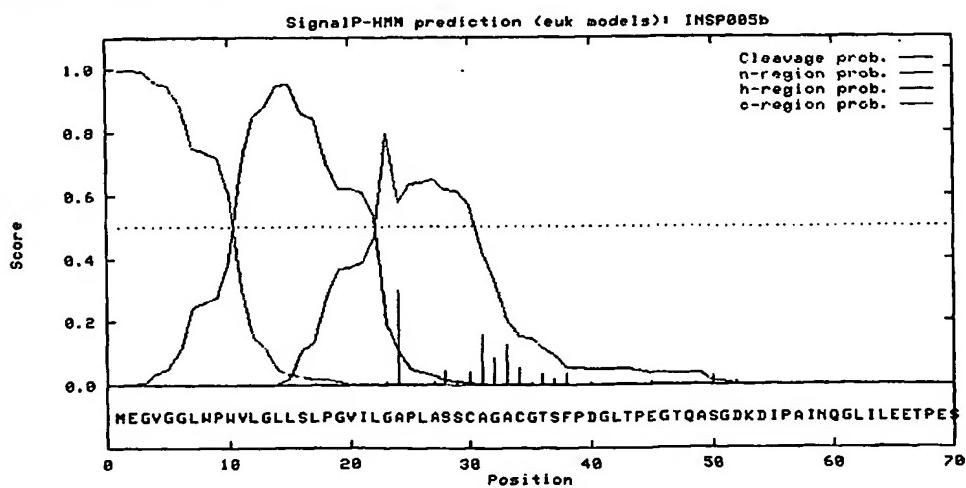
SignalP-NN result:



data

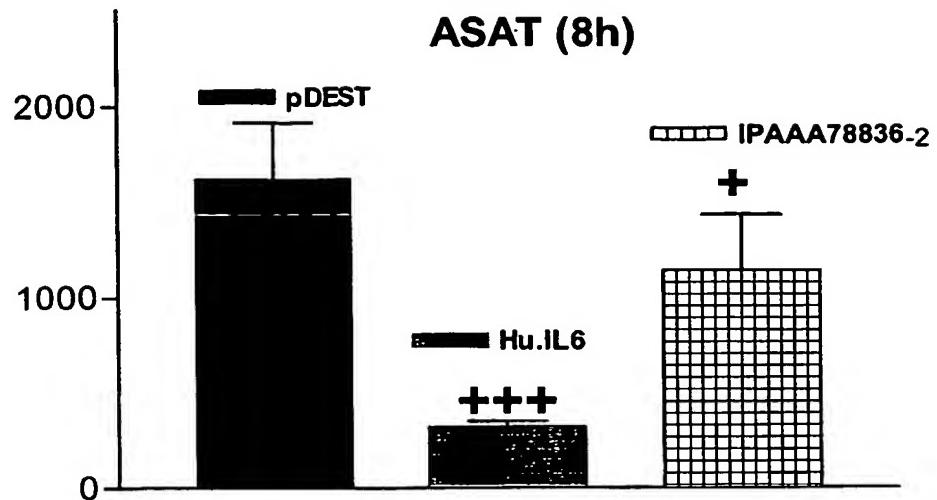
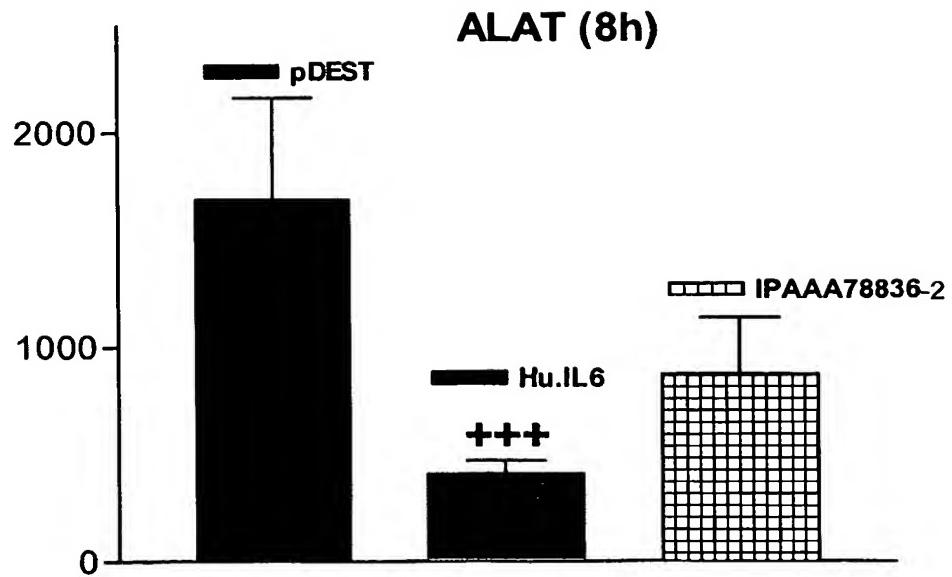
```
>INSP005b      length = 70
# Measure  Position  Value  Cutoff  signal peptide?
max. C    24      1.000  0.33   YES
max. Y    24      0.789  0.32   YES
max. S    13      0.991  0.82   YES
mean S   1-23     0.929  0.47   YES
# Most likely cleavage site between pos. 23 and 24: ILG-AP
```

SignalP-HMM result:

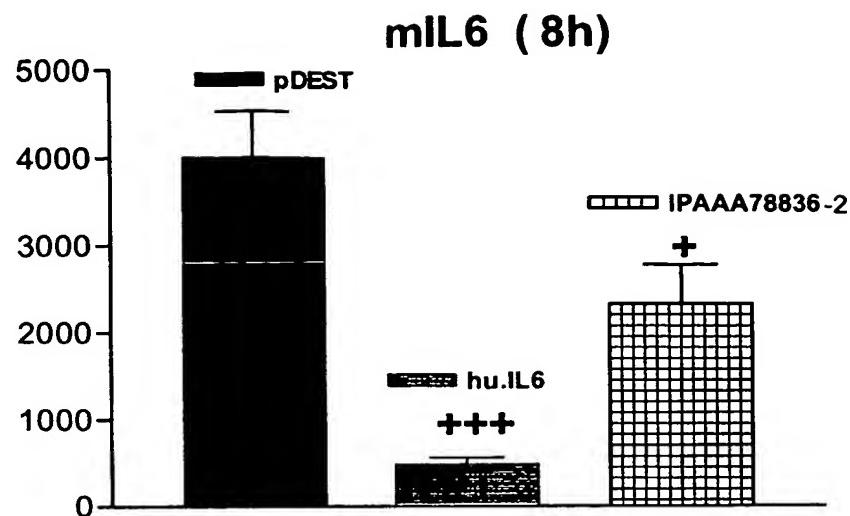


data

```
>INSP005b
Prediction: Signal peptide
Signal peptide probability: 0.996
Signal anchor probability: 0.003
Max cleavage site probability: 0.302 between pos. 23 and 24
```

FIGURE 15A**FIGURE 15B**

18/18

FIGURE 16A**FIGURE 16B**